◎ 논 문

http://dx.doi.org/10.22680/kasa2022.14.3.071

차량 안전 제어를 위한 파티클 필터 기반의 강건한 다중 인체 3차원 자세 추정

박준상* · 박형욱**

Particle Filter Based Robust Multi-Human 3D Pose Estimation for Vehicle Safety Control

Joonsang Park*, Hyungwook Park**

Key Words: Human Pose Estimation(인체 자세 추정), Sensor Fusion(센서 융합), Convolutional Neural Network(합성곱 신경망), Driver Monitoring System(운전자 모니터링 시스템)

ABSTRACT

In autonomous driving cars, 3D pose estimation can be one of the effective methods to enhance safety control for OOP (Out of Position) passengers. There have been many studies on human pose estimation using a camera. Previous methods, however, have limitations in automotive applications. Due to unexplainable failures, CNN methods are unreliable, and other methods perform poorly. This paper proposes robust real-time multi-human 3D pose estimation architecture in vehicle using monocular RGB camera. Using particle filter, our approach integrates CNN 2D/3D pose measurements with available information in vehicle. Computer simulations were performed to confirm the accuracy and robustness of the proposed algorithm.

1. 서 론

지금까지 차량 안전 제어는 주로 법규에서 규제하는 특정 자세를 기준으로 승객이 차량에서 정면을 향한 자세로 똑바로 앉는 것을 가정하여 연구되었다. 그러나 자율주행 환경에서 승객은 다양한 자세로 승차할 수 있게 되었으며 이러한 OOP(Out Of Position) 승객의 안전을 위해 2D 카메라, ToF 카메라, 실내 RADAR 등을 이용한 3D 자세 추정을 활용할 수 있다. 이 중 가장 저렴한 방법은 2D 카메라를 사용하는 것이지만, 인체 감지, 인체 2D 자세 추정, 인체 3D 자세 추정과 같은 복잡한 이미지 처리 알고리즘을 사용해야 한다.(1)

전통적인 컴퓨터 비전 기반 인체 감지 방법이 많이 연구되어 왔지만 최근 가장 일반적으로 사용되는 방법은 합성곱 신경망(CNN: Convolutional Neural Network)이다. CNN은 인체를 감지할 수 있을 뿐만 아니라 인체의 자세 추정을 가능하게 하는 해부학적 키포인트를 추출할 수 있다. 각 사람에 대해 1인 자세 추정을 수행하는 하향식 방법(2,3) 혹은 단일 탐지에 의해 신체 부위가 추출되고 개별 사람과 연결되는 상향식 방법(4,5)으로 구현된다.

또한 많은 연구자들이 사람의 3D 자세 추정 문제를 해결하려고 시도했다. 그래프 신경망(GNN: Graph Neural Network)과 인체 3D 자세 데이터 증강기법 등이 연구되어 왔으며 괄목할 만한 성취를 거두고 있다. 예시로 Human3.6m 데이터 세트에서 단일 이미지 활용 1인 3D 자세 추정을 수행하여 약 50~60mm 수준의 평균 관절 위치 오차(MPJPE: Mean Per Joint Position Error)를 달성했다.(6,7) 또한, MuPoTS-3D 데이터 세트에서 다중 인체 3D 자세 추정을 수행하여 약 70~80% 수준의 3DPCK를 달성했다.(8,9)

* 현대자동차, 연구원
** 현대자동차, 책임연구원
E-mail: rune2002@hyundai.com

논문접수일: 2022.5.27, 논문수정일: 2022.7.28, 게재확정일: 2022.8.25

차량 안전 제어 활용에 충분해 보일 수 있으나 딥러닝 기반 접근 방식은 설명할 수 없는 실패로 인해 잠재적인 위험이 따른다. 이러한 고질적인 한계를 극복하기 위해 자율주행 환경 내 설명 가능한 AI(XAI: Explainable AI)가 연구되어 왔다.[10,11] 다만, 이러한 접근 방식들은 AI 모델의 설명가능성(Explainability) 혹은 해석가능성(Interpretability)을 개선하거나 특정 출력이 어떤 패턴에 의해 활성화되어 나타났는지에 대한 연구가 대부분으로 기존 룰 베이스 알고리즘과 다르게 사람은 AI 모델을 완전히 이해하기 힘들다.

이러한 맥락에서 우리는 AI 모델이 실패하는 상황을 고려한 차량 내 실시간 다중 인체 3D 자세 추정 알고리즘을 제안한다. 본 연구는 파티클 필터를 사용하여 최대 사후 확률 추정을 통해 차량에서 사용 가능한 정보와 CNN 2D/3D 인체 자세 측정을 통합한다. 먼저 파티클 필터 알고리즘 수행을 위해 승객 추적 알고리즘을 사용하여 CNN으로 감지된 승객과 기존 승객을 매칭한다. 그 후, 파티클은 IMU 센서와 같은 동역학 센서 및 인체 동역학 모델을 통해 전파되며 파티클의 가중치는 사전 정보(예: 불가능한 자세 및 차량 내 제약)와 CNN 측정값을 사용하여 업데이트된다. 최종적으로 전파된 파티클과 업데이트된 가중치를 통해 승객의 3D 자세를 추정한다.

논문의 나머지 부분은 다음과 같이 구성된다. 2장에서는 CNN을 사용한 키포인트 추출 및 승객 추적 알고리즘을 제시한다. 3장에서는 파티클 필터를 통한 상태 추정 방법을 제시하고 4장에서는 실험 결과를 제시한다. 5장에서는 논문을 마무리하고 향후 연구 방향을 제시한다.

2. 키포인트 추출 및 승객 매칭

2.1. 키포인트 추출

본 연구에서는 CNN을 활용하여 승객의 2D 및 3D 키포인트를 추출한다. 추출된 키포인트는 Fig. 1 왼쪽과 같이 트리 구조이며 E개의 간선을 갖는다. 여러 SOTA 모델 중에서 경량화된 OpenPose[12]를 기반으로 Occlusion-Robust Pose-Maps[8]를 구현한 모델[13]을 선정했다. 선정 모델은 MobileNet을 백본으로 경량화되어 실시간 자세 추정에 용이하며 가림 현상에도 강건하다. 또한, 해당 모델은 상향식 방법으로 구현되어 Cao et al.이 주장하는 바와 같이 다수의 승객 검출에도 비슷한 계산시간이 소요된다.[4]

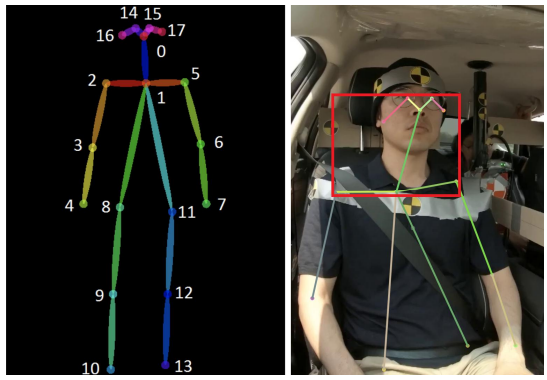

Fig. 1 Configuration of keypoints (left) and an example of detected keypoints and a generated bounding box including head, shoulders, and neck (right)

2.2. 승객 추적 알고리즘

파티클 필터와 같은 순차 상태 추정을 적용하기 위해 CNN이 새로 감지한 승객을 기존 승객에 매칭할 필요가 있다. 우리는 헝가리안 알고리즘을 활용하여 Fig. 2와 같이 새로 감지한 승객의 경계박스와 기존 승객의 경계박스 간의 IoU(Intersection Over Union) 합을 최대화하는 방식으로 구현했다. 기존 방식[14]과 달리 승객의 경계박스는 Fig. 1 오른쪽과 같이 승객의 관심 키포인트들을 모두 포함하는 가장 작은 직사각형으로 생성된다. 기존 승객은 예측된 3D 좌표를 이미지 평면에 정사영한 키포인트를 사용한다. 새로 감지된 승객이 매칭되지 않으면 새로운 승객으로 인식하며 기존 승객이 일정 횟수 이상 매칭되지 않으면 사라진 것으로 판단하여 삭제한다.

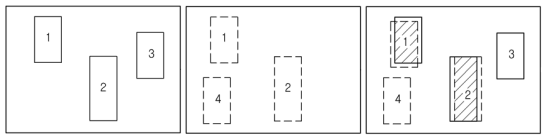

Fig. 2 An example of the passenger tracking algorithm, bounding boxes of detected passengers by CNN (left), those of existing passengers (middle), and a matching solution (right)

3. 파티클 필터 활용 상태 추정

3.1. 시스템 상태 및 파티클 정의

각 파티클은 추정하고자 하는 인체 모델의 매개변수 θ

와 가중치 w로 구성된다. 인체 모델의 매개변수 θ는 식 (1)과 같이 구성된다.

$$\theta = [x_0 \;\; \dot{x_0} \;\; \alpha \;\; \dot{\alpha} \;\; \beta \;\; \dot{\beta} \;\; l]^T \tag{1}$$

여기서, x_0는 루트 키포인트의 카메라 좌표계에 대한 3차원 위치 벡터, $l = \{l_i\}_{i=1}^{E}$는 신체 길이, $\alpha = \{\alpha_i\}_{i=1}^{E}$, $\beta = \{\beta_i\}_{i=1}^{E}$는 관절 각도이다. 이 때 루트 키포인트는 가장 적게 가려지는 키포인트(목 또는 골반)를 권장하며 본 논문에서는 목 키포인트를 사용했다. 신체 길이는 각 키포인트 사이의 거리를 의미한다. j번째 키포인트 3D 좌표 x_j는 식 (2)와 같이 구할 수 있다.

$$x_j = x_0 + \sum_{i}^{i \in p(j)} (l_i [\sin\alpha_i \cos\beta_i \;\; \sin\alpha_i \sin\beta_i \;\; \cos\alpha_i]^T) \tag{2}$$

여기서, $p(j)$는 루트 키포인트에서 j번째 키포인트까지의 경로에 포함되는 간선의 집합이다. 반대로 3D 좌표로부터 x_0, α, β, l을 구할 수 있다.

3.2. 파티클 초기화

2.2절의 승객 추적 알고리즘에 의해 승객이 새로 감지되면 하나의 파티클 필터가 생성된다. 초기 CNN 3D 측정값으로부터 계산된 매개변수와 초기화 노이즈 벡터 $n_{\in}$의 합으로 N개의 파티클이 초기화된다. 이 때 $n_{\in}$는 평균이 0인 정규 분포로 모델링되며 매개변수 계산 시 미분 성분들은 모두 0으로 가정한다.

3.3. 파티클 전파

파티클은 식 (3)과 같이 IMU 센서와 같은 차량 동역학 센서를 사용하여 동역학 모델에 의해 전파된다.

$$\theta_{k|k-1} = f(\theta_{k-1|k-1}, a, w, dt) + n_k \tag{3}$$

여기서, $\theta_{k|k-1}$은 k스텝 사전 추정, $\theta_{k-1|k-1}$은 $k-1$스텝 사후 추정, a, w는 각각 동역학 센서의 가속도계 및 각가속도계 측정값, dt는 미소 시간, f는 동역학 모델, n_k는 평균이 0인 정규 분포로 모델링된 k스텝 전파 노이즈 벡터이다. 동역학 모델은 높은 가속도 또는 각가속도의 동적인 운전 상황에서 승객의 급격한 움직임으로 인해 파티클 필터가 승객을 놓치는 것을 방지한다. 본 논문에서는 식 (4)와 같이 루트 키포인트가 a만큼 가속되는 간단한 모델을 사용했다.

$$\theta_{k|k-1} = A\theta_{k-1|k-1} + Ba + n_k$$
$$A = \begin{bmatrix} P_3 & & & \\ & P_E & & \\ & & P_E & \\ & & & I_E \end{bmatrix}, \; B = \begin{bmatrix} Q_3 \\ \\ \\ \end{bmatrix}$$
$$P_m = \begin{bmatrix} 1 & dt \\ 0 & 1 \end{bmatrix} \otimes I_m, \quad Q_m = \begin{bmatrix} 0 \\ dt \end{bmatrix} \otimes I_m \tag{4}$$

여기서, $\otimes$는 크로네커 곱, I_m은 $m \times m$ 단위행렬, 공백은 영행렬이다.

3.4. 가중치 업데이트

3.4.1. 사전 정보를 통한 가중치 업데이트

파티클의 가중치는 여러 사전 정보를 사용하여 업데이트될 수 있다. 승객이 신체적으로 불가능한 자세를 취하고 있거나 카메라의 화각 밖 또는 차량 외부에 있는 경우 파티클의 가중치를 낮출 수 있다. 차량의 시트 착좌 센서를 통해 승객이 시트에 착좌하고 있음을 알 수 있는 경우, 예상되는 키포인트의 3D 좌표와 비교해 가중치를 업데이트할 수 있다.

3.4.2. CNN 측정값을 통한 가중치 업데이트

새로 감지된 승객이 2.2 절의 승객 추적 알고리즘에 의해 기존 승객에 매칭되는 경우 해당 CNN 측정값을 통해 가중치 업데이트가 수행된다. 3.1절에서 언급했듯이 키포인트의 3D 좌표는 파티클 매개변수 θ로부터 계산할 수 있다. 파티클로부터 계산된 3D 좌표와 CNN 3D 측정값을 비교한다. 또한 키포인트를 픽셀 이미지 평면에 대해 투영한 좌표와 CNN 2D 측정값을 비교한다. CNN 2D 및 3D 측정은 서로 독립인 다변수 정규 분포로 모델링되며 i번째 파티클의 가중치는 식 (5)에 의해 업데이트된다.

$$w_k^i = w_{k-1}^i P(D|\theta_i) P(X|\theta_i) \tag{5}$$

여기서, w_k^i는 k스텝 i번째 파티클의 가중치이고 D, X는 각각 CNN 2D, 3D 측정값이다.

3.5. 최대 사후 확률 추정

현재 상태 벡터가 최대 사후 확률(MAP: Most A Posteriori) 추정이기도 한 칼만 필터와 달리 파티클 필터에는 가중치가 있는 N개의 서로 다른 후보가 있다. MAP 추정은 모든 파티클의 가중 평균을 계산하여 얻는다. 이는 단순히 가장 높은 가중치를 가진 파티클을 선택하는 것보다 더 강건하기 때문이다. 가중 평균을 계산하기 전에 식 (6)과 같이 모든 가중치의 합이 1이 되도록 정규화해야 한다.

$$\theta_{k|k} = \sum_{i=1}^{N} w_k^i \theta_i, \ \sum_{i=1}^{N} w_k^i = 1 \tag{6}$$

3.6. 파티클 리샘플링

파티클 필터의 퇴화(Degeneracy) 문제를 방지하기 위해 유효 파티클 수의 추정치 $\widehat{N_{eff}}$가 작거나[15] 사전 정보를 통해 유효하지 않다고 판단된 파티클이 임계값 이상 존재할 때 파티클을 리샘플링한다. 이 때 $\widehat{N_{eff}}$는 식 (7)에 의해 계산된다.

$$\widehat{N_{eff}} = \frac{1}{\sum_{i=1}^{N} (w_k^i)^2} \tag{7}$$

여기서, w_k^i는 식 (6)에 의해 정규화된 가중치이다.

4. 실 험

4.1. 실험 환경

실험은 Fig. 3과 같이 차량 앞유리에 고정된 Intel RealSense D455 깊이 카메라가 장착된 PC에서 수행되었고 깊이 카메라에 Bosch BMI055 IMU 센서가 내장되어 있다. 제안된 알고리즘은 카메라의 RGB 이미지와 IMU 센서를 활용한다. CNN과 같은 영상 처리가 필요한 구성 요소들은 100ms 주기로 실행되고 다른 구성 요소들은 20ms 주기로 실행된다.

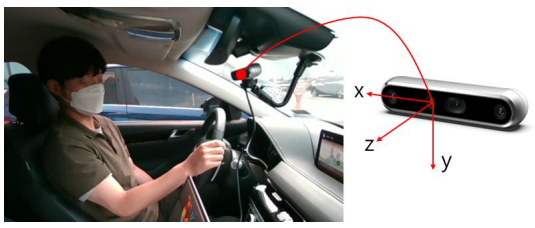

Fig. 3 Configuration of RealSense D455

4.2. 시나리오

4.2.1. 시나리오 1: FCA 급제동

차량이 전방 충돌 방지 보조(FCA: Forward Collision-avoidance Assist) 시스템에 의해 급제동하는 상황(Fig. 4 참조)을 모사했으며 종방향으로 급속한 승객 거동을 보인다.

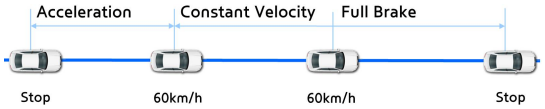

Fig. 4 Test Protocol of Scenario 1

4.2.2. 시나리오 2: 더블 레인 체인지

차량이 높은 속도에서 연속으로 차선을 두 번 바꾸는 상황(Fig. 5 참조)을 모사했으며 횡방향으로 급속한 승객 거동을 보인다.

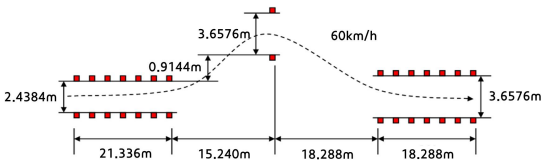

Fig. 5 Test Protocol of Scenario 2

4.3. 실험 결과

상해 주요 포인트(머리, 목, 오른쪽 어깨, 왼쪽 어깨)의 카메라 좌표계에 대한 3D 좌표 추정값과 Intel RealSense D455 깊이 카메라 측정값(Ground Truth로 가정)을 비교하였고 그 결과 중 가장 큰 움직임을 보이는 시나리오 1에서 동승석 머리 Z 좌표와 시나리오 2에서 동승석 머리 X 좌표 결과를 Fig. 6에 정리하였다. 승객의 급격한 움직임으로 인해 CNN이 실패한 상황에서도 강건하게 승객의 3D 자세를 추정함을 확인할 수 있다.

정량적인 성능 비교를 위해 깊이 카메라 측정값에 대한

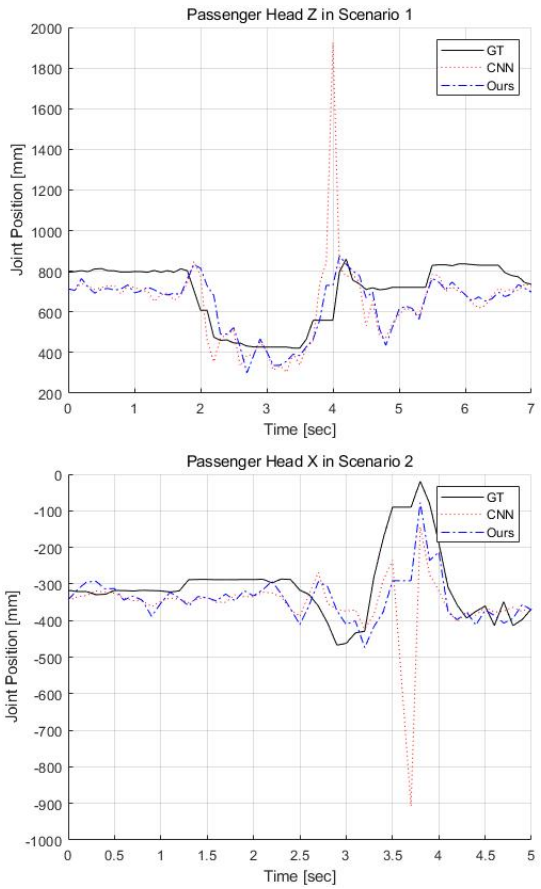

Fig. 6 Simulation results of 3D pose estimations

Table 1 Results on each scenario (A-MPJPE in mm)

Test Case		CNN	Ours
Scenario 1	1	274	218
	2	287	208
Scenario 2	1	233	212
	2	265	209

오차를 비교했다. 이 때 우리는 루트 키포인트가 원점인 상대 좌표를 비교하는 MPJPE가 아닌 카메라 좌표계에 대한 3D 좌표 오차를 활용한 A-MPJPE(Absolute Mean per Joint Position Error)[16]를 계산했다. Table 1은 각 시나리오에 대한 제안 알고리즘과 알고리즘을 적용하지 않은 CNN의 결과를 보여준다. 제안 알고리즘은 A-MPJPE가 낮아 가혹한 운전 환경에서 3D 자세 추정 성능을 향상시킴을 알 수 있다.

5. 결 론

이 연구에서 차량 안전 제어에 활용할 수 있도록 파티클 필터를 사용하여 차량 내 강건한 실시간 다중 인체 3D 자세 추정 알고리즘을 제안했다. 본 알고리즘은 CNN과 차량 센서를 성공적으로 통합했으며 실험을 통해 CNN이 실패하는 동적인 운전 상황에서도 정확성과 강건함을 보여줌을 확인했다. 이 논문은 OOP 승객을 위한 자세 감지 및 차량 안전 제어 연구를 위한 좋은 출발점을 제공한다.

참고문헌

(1) X. Ji, Q. Fang, J. Dong, Q. Shuai, W. Jiang and X. Zhou, 2020, "A survey on monocular 3D human pose estimation," Virtual Reality & Intelligent Hardware, Vol. 2, No. 6, pp. 471~500.

(2) K. Sun, B. Xiao, D. Liu and J. Wang, 2019, "Deep High-Resolution Representation Learning for Human Pose Estimation," IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5686~5696.

(3) H. Fang, S. Xie, Y. Tai and C. Lu, 2017, "RMPE: Regional Multi-Person Pose Estimation," IEEE International Conference on Computer Vision (ICCV), pp. 2334~2343.

(4) Z. Cao, T. Simon, S. Wei and Y. Sheikh, 2017, "Realtime Multi-person 2D Pose Estimation Using Part Affinity Fields," IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1302~1310.

(5) M. Kocabas, S. Karagoz, and E. Akbas, 2018, "Multiposenet: Fast multi-person pose estimation using pose residual network," European Conference on Computer Vision (ECCV), pp. 437~453.

(6) L. Zhao, X. Peng, Y. Tian, M. Kapadia and D. N. Metaxas, 2019, "Semantic Graph Convolutional Networks for 3D Human Pose Regression," IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3420~3430.

(7) K. Gong, J. Zhang and J. Feng, 2021, "PoseAug: A Differentiable Pose Augmentation Framework for 3D Human Pose Estimation," IEEE Conference on Computer Vision and Pattern Recognition

(CVPR), pp. 8571~8580.
(8) D. Mehta et al., 2018, “Single-Shot Multi-person 3D Pose Estimation from Monocular RGB,” International Conference on 3D Vision (3DV), pp. 120~130.
(9) J. Lin, G. H. Lee, 2020, “HDNet: Human Depth Estimation for Multi-person Camera-Space Localization,” European Conference on Computer Vision (ECCV), pp. 633~648.
(10) J. Kim, S. Moon, A. Rohrbach, T. Darrell and J. Canny, 2020, “Advisable Learning for Self-Driving Vehicles by Internalizing Observation-to-Action Rules,” IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 9661~9670.
(11) Y. Xu, X. Yang, L. Gong, H. Lin, T. Wu, Y. Li and N. Vasconcelos, 2020, “Explainable Object-Induced Action Decision for Autonomous Vehicles,” IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 9523~9532.
(12) D. Osokin, 2018, “Real-time 2D Multi-Person Pose Estimation on CPU: Lightweight OpenPose,” arXiv:1811.12004.
(13) D. Osokin, 2020, Real-time 3D Multi-person Pose Estimation Demo [Source code]. https://github.com/Daniil-Osokin/lightweight-human-pose-estimation-3d-demo.pytorch.
(14) A. Bewley, Z. Ge, L. Ott, F. Ramos and B. Upcroft, 2016, “Simple online and realtime tracking,” IEEE International Conference on Image Processing (ICIP), pp. 3464~3468.
(15) M. S. Arulampalam, S. Maskell, N. Gordon and T. Clapp, 2002, “A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking,” IEEE Transactions on Signal Processing, Vol. 50, No. 2, pp. 174~188.
(16) M. Veges and A. Lorincz, 2019, “Absolute Human Pose Estimation with Depth Prediction Network,” International Joint Conference on Neural Networks (IJCNN), pp. 1~7.